CAST: Interim Report to the ASTB, May 19th 1999

Ben Whitaker

University of Leeds

Introduction

Collaboratory *for*
Atmospheric
Science *and*
Technology

Interim Report to ASTB

May 19 1999

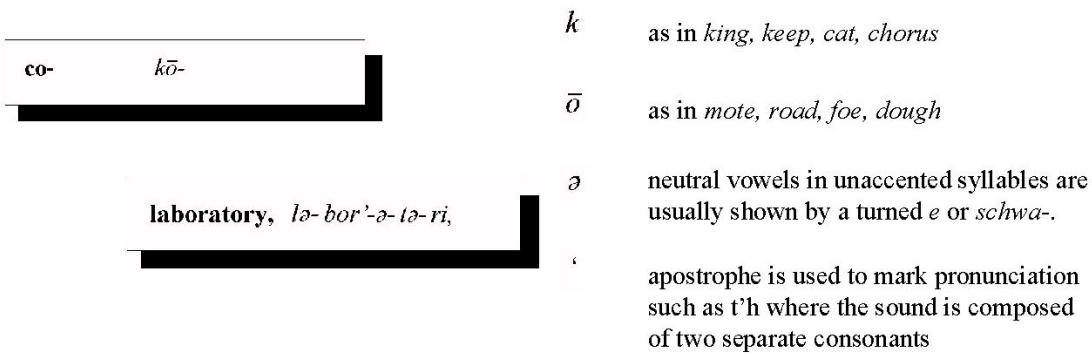
http://cast.nerc.ac.uk/

The broad aim of the CAST project is to promote interaction and collaboration between UK atmospheric scientists. The project provides a platform for exploring the use of networked based techniques and tools not only to improve the ease of collaboration but also to develop new methods of remote working to enhance the research environment for atmospheric scientists. The trial systems we have been working on are more limited than eventual applications, where interactions between modellers, field and laboratory scientists will be required, but they are designed to provide an adequate proof of concept.

Electronic networks to support Atmospheric Science already exist:

- UGAMP - mainly computational
- UWERN - Universities' Weather Research Network (mainly computational but includes field work and modelling)
- CAST - aimed initially at the type of work done in, for example, ACSOE (chemistry/physics/field/modelling) where there is a greater diversity of user. Involves BADC, Cambridge and Leeds

Connecting Information



A model for data integration is the Collaboratory . We originally adopted the spelling of Colaboratory with a single l to help users to adopt the pronuncation

Co-laboratory

instead of the more difficult

Col-lab-ut-ri

or

Col-lab-brut-tri

which seems to trip speakers up. However, the more common and popular spelling is "Collaboratory". The idea of the Collaboratory is very close to Berners-Lee's original conception of the WWW as a linked informmation resource, but as the Internet technology has developed a greater degree of interaction is now possible.

Traditional models of Collaboration

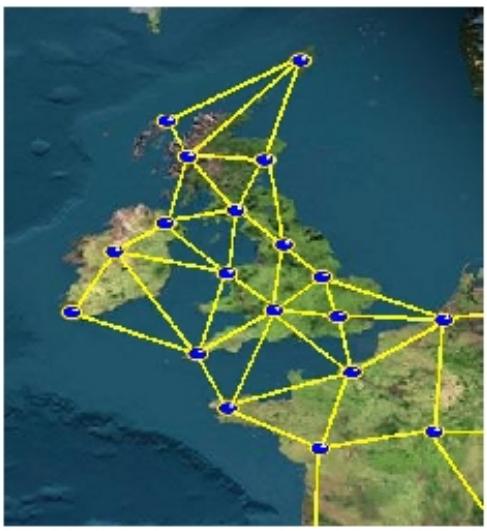
- meetings
- telephone
- email
- conferences
- publication

In the past, collaboration has manifest itself in two ways:

Personal interaction:
discussion and exchange of ideas required to carry out research programs. In the pre-electronic era such collaboration was mainly achieved via face-to-face meetings, telephone conversations, and the exchange of paper documents by post. Although today the use of electronic mail is widespread, personal interaction, particularly in the early and late reporting stages of a research project, is usually made via presentation and discussion in a seminar setting.

Dissemination:
usually through the publishing of work in scientific journals or at conferences. Here the aim is to inform the wider scientific community of progress. Less attention has been paid to the public dissemination of science, although this function is fulfilled by the research councils and learned societies.

The Collaboratory model

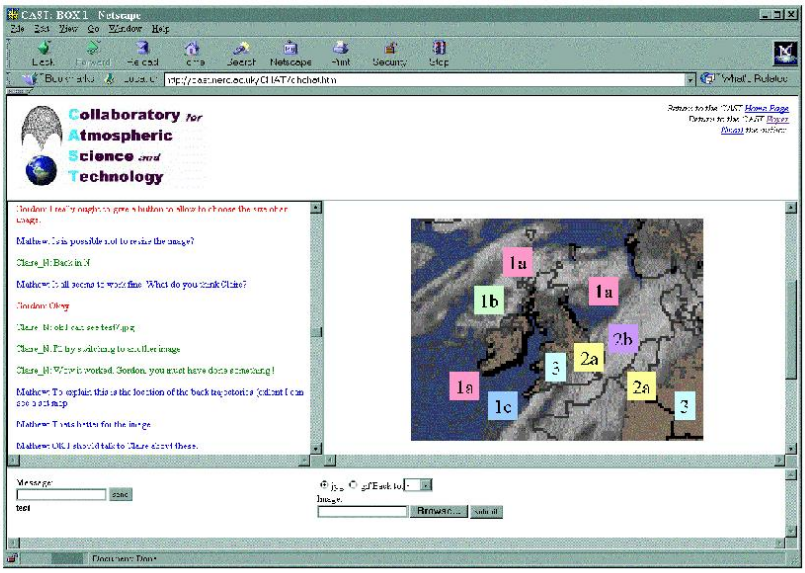


- enhancement not replacement
- electronic tools for
  - data sharing
  - data visualisation
  - group working

The CAST project aims to provide communication links between geographically remote groups, and between the research community and funding bodies, commercial organisations, and the public. These are not, however, intended to replace existing means of communication.

The CAST project seeks to develop resources which are only possible using Internet based tools. Such resources encompass the manipulation and transmission of electronic media, and provide access to computational resources such as data archives and application programmes that may not be available locally.

The virtual meeting room



Web-based interaction can be broadly split into two modes

- Interaction between humans.
- Interaction between humans and machines.

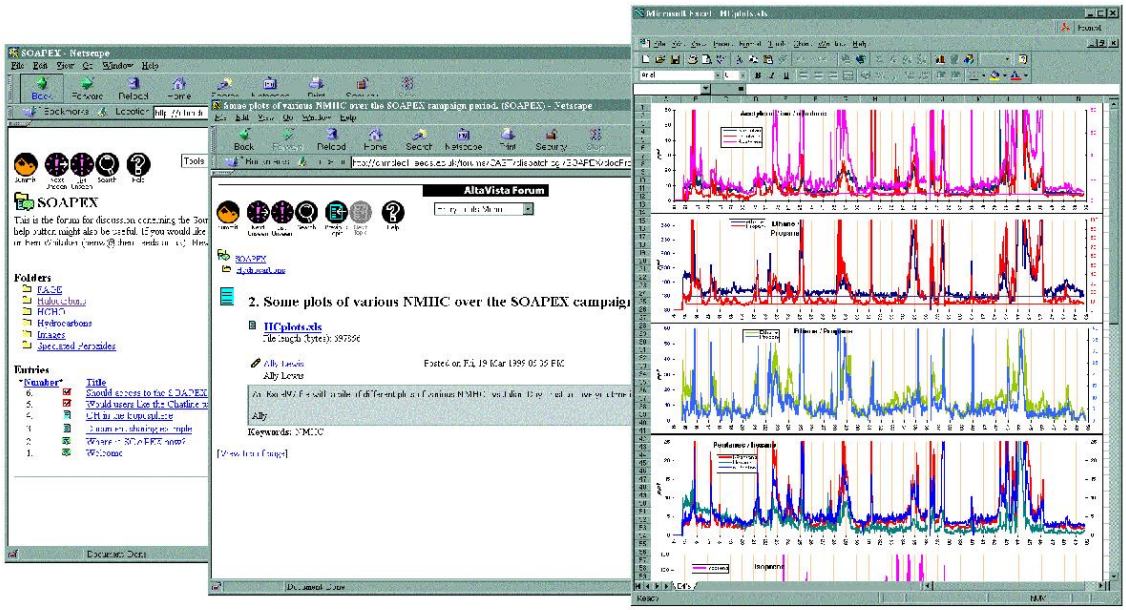
Each mode of interaction can be split into a further two groups.

Synchronous Interaction
The real time interaction between researchers observed when people talk face-to face, on the ‘phone, or an electronic chat-room. Synchronous interaction also includes the interaction between humans and fast machine resources which generate data rapidly.

Asynchronous Interaction
The exchange of ideas and resources can be achieved in a more disjoint manner by leaving data in archives or caches for retrieval by other users later. Perhaps the most widely used asynchronous medium is email. However, the CAST can provide a number of additional methods of exchanging data in an asynchronous manner.

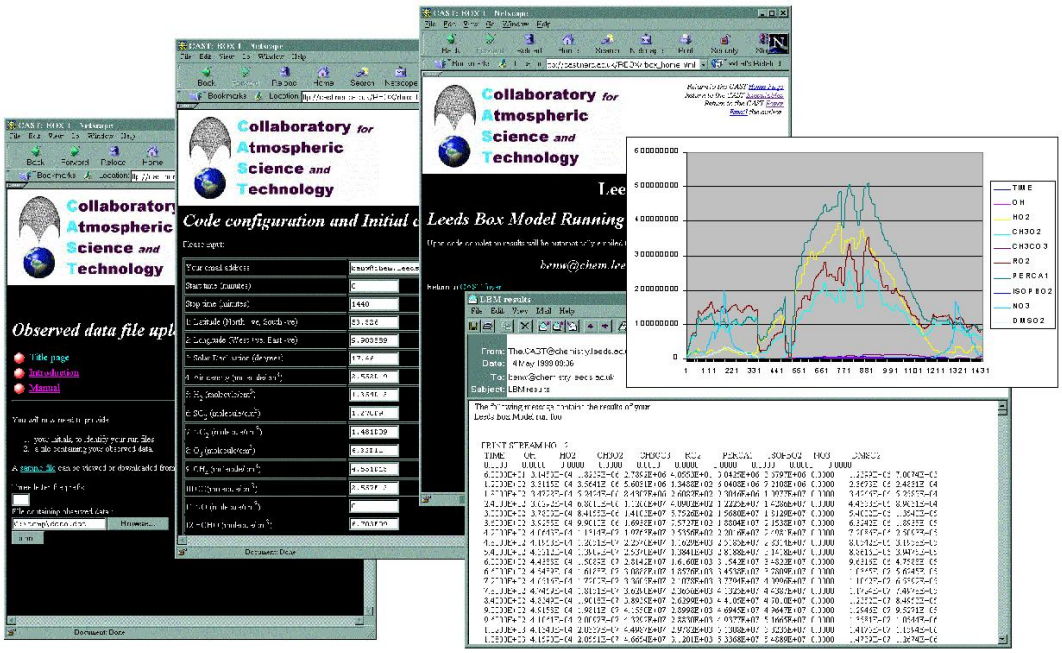
The screen shot shows a real time conversation between Gordon Hart (Leeds), Clare Reeves (Norwich) and Matt Evans (Cambridge). They are discussing the MAXOX 1999 flight plans.

Discussion Forums



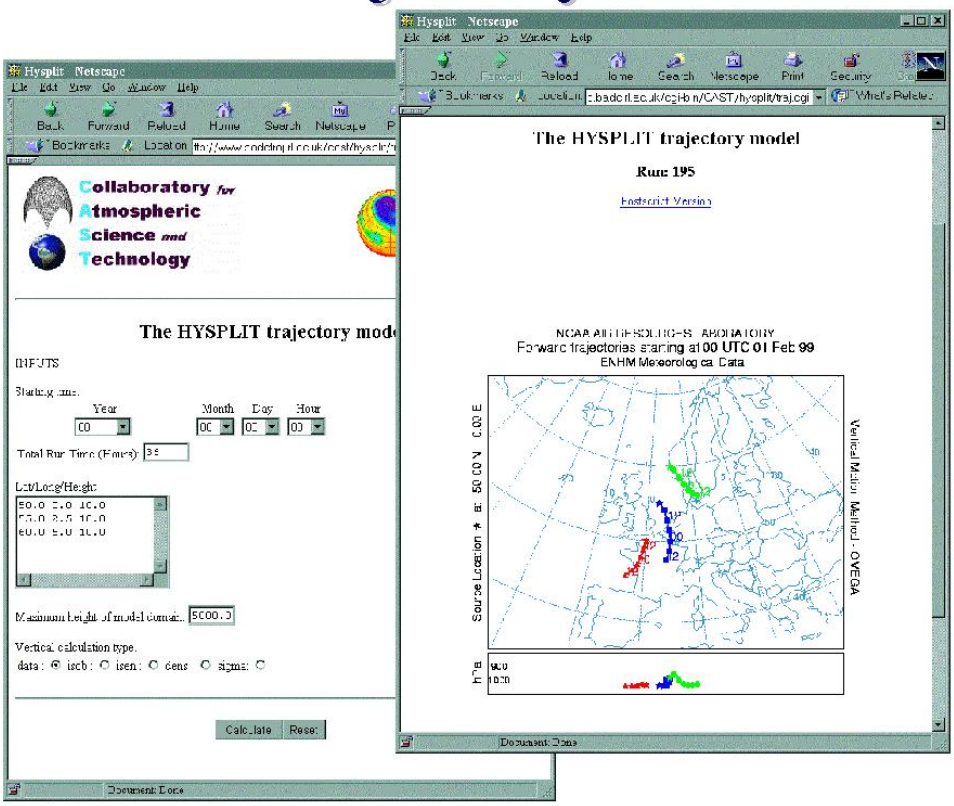
The Discussion Forum on the CAST is an example of asynchronous interaction. Here registered users can exchange and develop ideas through a “bulletin board” metaphor. Documents (Word, eXcel etc.) can be left on the site or in folders that users can create. Documents can be annotated, picked up and edited by other users of the forum. The slide illustrates this with a few screen shots of the SOAPEX campaign forum. Data for hydrocarbon measurements over the campaign are shown in the form of a eXcel spreadsheet which has been uploaded into a folder on the site. Users can set-up their account so that they are notified when new documents appear in any particular area of the site, or when existing documents are modified. Members of a forum can set themselves up in “teams” to work collaboratively on sub-themes of the overall project. The forum can be administered by any user with the necessary privileges remotely.

Gateways to applications



An example of asynchronous interaction between human and machine is the CAST gateway to the Leeds box model. Here a user (perhaps in the field) can upload a file containing observational data to the Leeds photochemical box model. Some minutes to hours later, depending on the complexity of the calculation requested, they receive by email a file containing the predicted key radical concentrations which can easily be read into a visualisation programme such as eXcel.

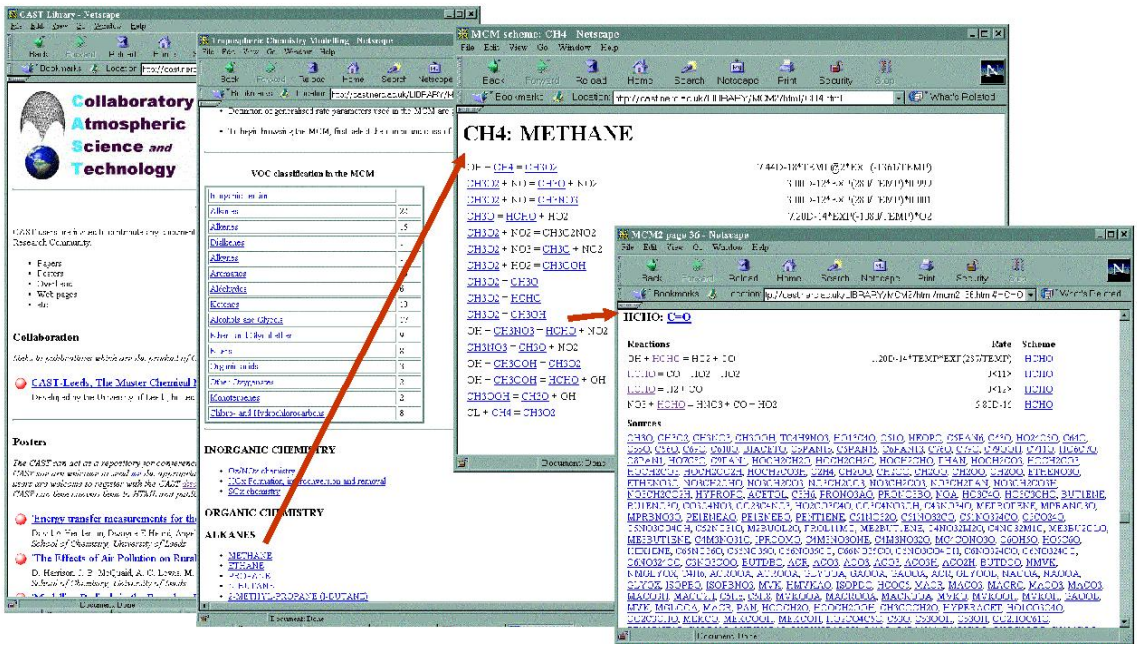
Trajectory viewer



An example of synchronous machine human interaction is the CAST trajectory viewer. It differs from the Cambridge trajectory server (UGAMP/ACSOE) in that the trajectories are calculated “on-the-fly” from ECMWF data held at BADC. The service makes use of the HYbrid Single-Particle Lagrangian Trajectory (HYSPLIT) program written by Roland Draxler (NOAA).

Input data are read from a form and submitted to the HYSPLIT running on a machine at BADC. HYSPLIT uses additional input from the ECMWF data set stored at BADC to return the trajectories requested to the clients browser window. A better quality postscript version of the trajectories can also be downloaded if required. It is envisaged that this could be a useful tool for field workers in the early stages of data interpretation (while still in the field)

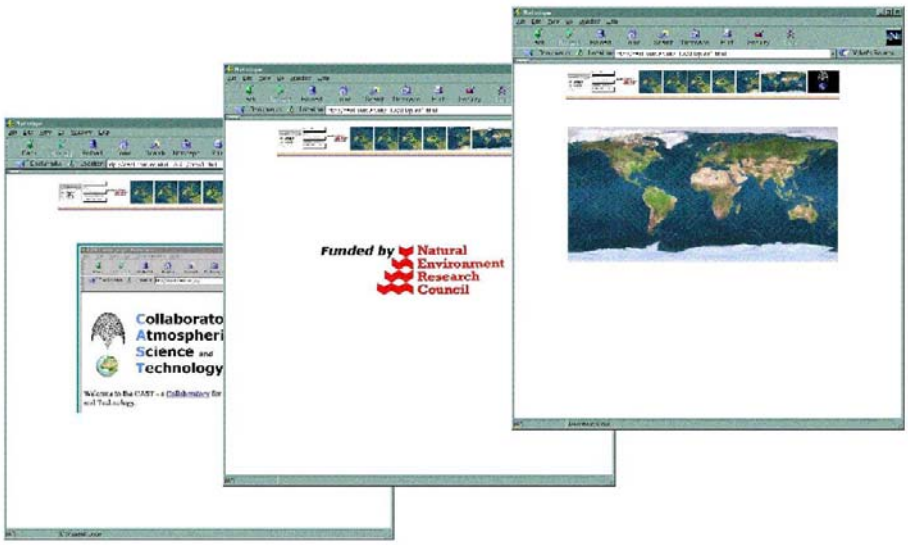
Dissemination



A web “hosting” service is provided by the CAST. An example is “Poster Heaven” - a repository for old conference posters (much appreciated by the post graduate students).

On-line publication of material is not confined to the delivery of static documents describing completed research. The Web is an excellent method for the dissemination of dynamic documents which become updated as data is collected and collated, discussion proceeds, knowledge expands and opinions change. An example is the Master Chemical Mechanism that describes the oxidation chemistry of the major VOCs..

Public understanding of science



The CAST RealServer and ShockWave services allow users to leave pre-recorded lecture presentations on the CAST server to be delivered at the request of users. Once running properly users can leave anything from single presentations, to seminars to entire mutimedia lecture courses with the CAST.

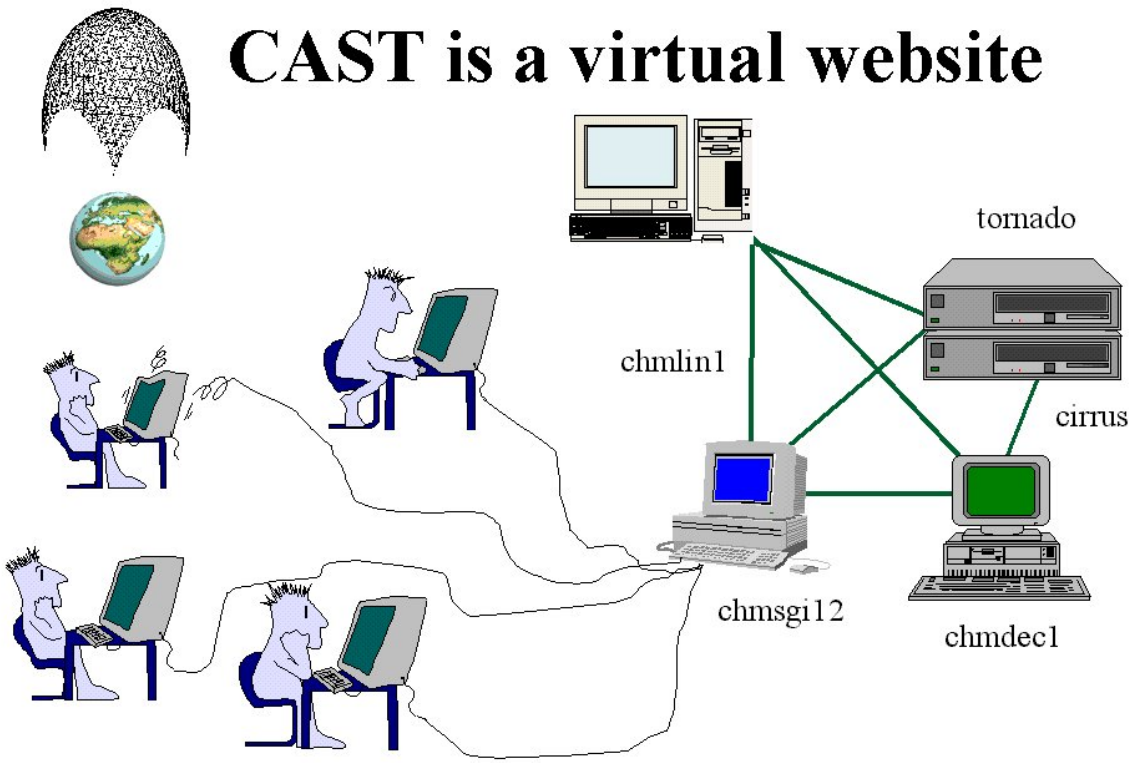
Quo vadis?

- work with the community
  - publicity
  - usability
  - security
- field work trials
- project exit strategy

CAST has been running for 7 months. Now that the basic technology is in place, in the last phase of the pilot project we need to work closely with the user community. Firstly to make them aware of what the CAST can provide; secondly to refine the user interface and make it more transparent to users; thirdly to ensure that access to information resources is secure (e.g. ECWMF data held at BADC or “pre-release” field campaign data held in the SOAPEX forum)

We would like to begin testing the collaboratory concept in a real campaign (hopefully PUMA)

If CAST is considered successful we need to consider the long term maintenance of the site



The CAST is not one resource at one site but many. Components of the CAST "server" run on 4 different machines at Leeds, 2 machines at BADC, and other in Cambridge.

The Alta Vista Discussion Forums can be managed by any authorised user located anywhere

CAST could provide a "collaboratory support officer" to NERC sponsored field campaigns